

# Searching Microblogs: Coping with Sparsity and Document Quality

**Nasir Naveed**, Thomas Gottron, Jérôme Kunegis, Arifah Che Alhadi  
**naveed**, gottron, kunegis, alhadi@uni-koblenz.de

Query: beer

Rank	Tweet
1	beer
2	beer!!
3	BEER
4	Beer
5	beer.....
6	To beer or not to beer on Beer Summit ?
7	beer beer beer beer beer beer beer. Simple 3pm
8	<a href="http://ping.fm/p/Bnra7">http://ping.fm/p/Bnra7</a> - In!!! BEER, BEER, BEER, BEER, BEER, BEER, BEER, BEER, BEER, BEER, BEER,
9	Lompoc. beer beer beer beer beer beer beer beer beer beer beer. <a href="http://twitpic.com/l68ld">http://twitpic.com/l68ld</a>
10	Beer beer beer beer beer beer beer beer beer beer beer beer beer beer beer. Er, guess what I'm looking forward to?

- Observation on Microblogs
  - ◆ Short tweets
  - ◆ Relevant (contain keywords), but not informative
  
- Hypotheses:
  - ◆ Length normalization in twitter is counter productive
  - ◆ Introduction of static quality measure of “*interestingness*” improves the retrieval results

- **Verbosity Hypothesis**

“Long document elaborates same topic longer, therefore, use same words”

- **Scope Hypothesis**

“Long document addresses several topics, therefore, use different words”

- ◆ Spearman’s rank correlation:

$\rho = 0.377$  (between tweet length and redundancy)

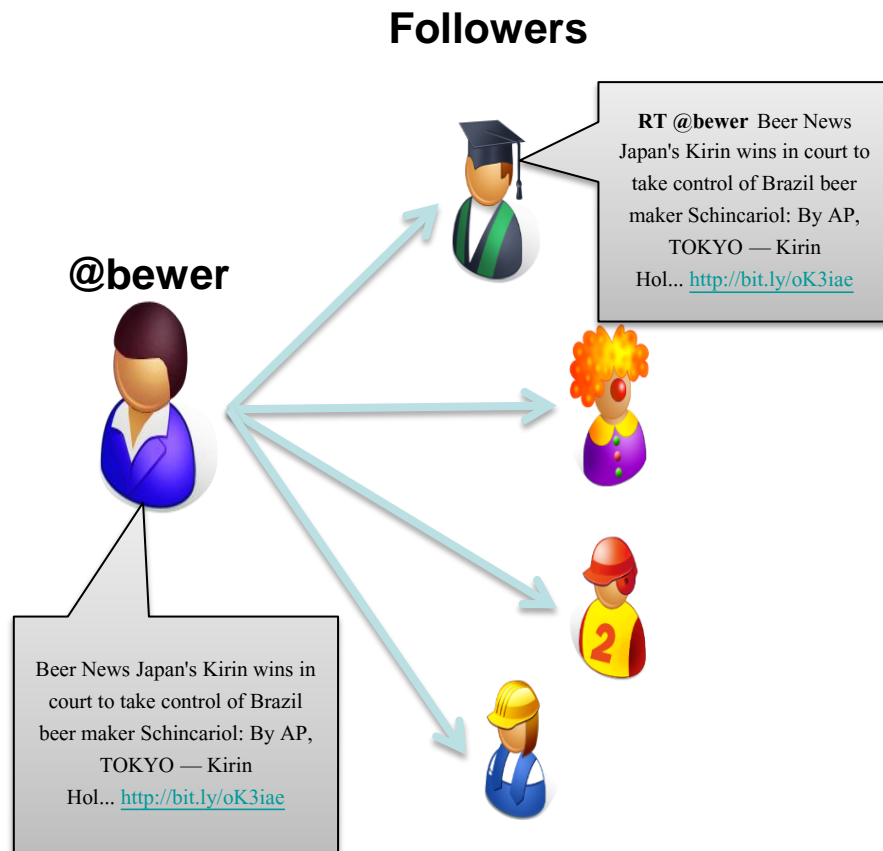
**Observation:**

<b>Tweets</b>	<b>No. of topics a tweet is related to</b>
77.1 %	1
99.6 %	2

**Tweets are not verbose**




**Tweets are focused**

**Length normalization:  
unjustified preference of short tweets**



- Retweet
  - ◆ *interesting* for wider audience
  - ◆ Sign of interestingnes
  - ◆ Sign of quality
- Depends on
  - ◆ Content
  - ◆ Social network
- Content based retweet prediction

**Retweet-Odds as sign of Quality**

Feature	Dimensions
Message feature	Direct message
	Username @Bob
	Hashtag #CIKM2011
	URL www.xyz.com
Sentiment	Valence 
	Arousal 
	Dominance 
Emoticons	Positive 😊
	Negative ☹️
Exclamation	Positive great
	Negative fail
Punctuation	!
	?
Terms	Odds
Topic	100 Topics

- Logistic regression

$$f(z) = \frac{1}{1 + e^{-z}}$$

$$z = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_nx_n$$

- Model parameters  $w_i$  learned on training data

Dataset	Users	Tweets	Retweets
Petrovic	4,050,944	21,477,484	8.46%

Feature	Dimensions	Weight
Constant	(intercept)	-5.45
Message feature	Direct message	-147.89
	Username	146.82
	Hashtag	42.27
	URL	249.09
Sentiment	Valence	-26.88
	Arousal	33.97
	Dominance	19.56
Emoticons	Positive	-21.8
	Negative	9.94
Punctuation	!	-16.85
	?	23.67
Terms	Odds	19.79

---

Topic	Weight
social media market post site web tool traffic network	27.54
follow thank twitter welcome hello check nice cool people	16.08
credit money market business rate economy home	15.25
christmas shop tree xmas present today wrap finish	2.87
home work hour long wait airport week flight head	-14.43
twitter update facebook account page set squidoo check	-14.43
cold snow warm today degree weather winter morning	-26.56
night sleep work morning time bed feel tired home	-75.19

---

- System setup

System	Properties
Lucene (baseline)	with length normalization
Lucene-noLen	with no length normalization
Retweet-odds	Lucene-noLen reranked using retweet odds

- Queries

- ◆ 20 queries (1 to 4 words of length)

- Evaluation strategies

- ◆ Relevance-based (MAP)
- ◆ Subjective evaluation

- Crowdsourcing (AMT)

- ◆ for relevance and system preference feedback

## Relevance Judgment

### Evaluate Search Results for Tweets

1. Please mark for EVERY tweet if it is RELEVANT or NOT RELEVANT and INFORMATIVE or NOT INFORMATIVE to the following query:

Input query: *beer*

Tweets	Relevant	Not relevant	Informative	Not informative
beer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lompoc. <a href="#">beer beer beer beer beer beer beer beer beer beer</a> . <a href="http://twitpic.com/168ld">http://twitpic.com/168ld</a>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
UK beer mag declares "the end of beer writing." <a href="#">@StanHieronymus</a> says not so in the US. <a href="http://bit.ly/424HRQ">http://bit.ly/424HRQ</a> #beer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Go green and turn those empty beer bottles into recycled beer glasses!   <a href="http://bit.ly/2src7F">http://bit.ly/2src7F</a> #beer #recycle (via: <a href="#">@td333</a> )	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
beer.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<a href="#">beer!!</a>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Team is just off to be 'usability tested' I wonder if they will be usable on their return?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
BEER^5. RT <a href="#">@dewbelle</a> : BEER BEER BEER BEER. RT <a href="#">@kulturbrille</a> : BEER BEER BEER. RT <a href="#">@Bluebarrow</a> : BEER BEER. RT <a href="#">@WalterMitty007</a> : BEER	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ooh I snagged a treasury! <a href="http://tinyurl.com/6j2gog">http://tinyurl.com/6j2gog</a> <a href="#">woot woot!</a>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## System Preference

### Compare Search Results for Tweets

#### Instructions

Imagine you are using a search engine for searching tweets.

In this task, we will provide you with two (2) search results for a given query. Your task is to look at these two resultsets and tell us which one you prefer and why?

#### 1. Which resultset do you prefer for the following query?

Input query: *beer*

1. UK beer mag declares "the end of beer writing." @StanHieronymus says not so in the US. <http://bit.ly/424HRQ> #beer

2. beer summit @bspward @jhinderaker no one had billy beer? heehee #narm - beer summit @bspward @jhinde <http://tinyurl.com/n29oxj>

3. Go green and turn those empty beer bottles into recycled beer glasses! | <http://bit.ly/2src7F> #beer #recycle (via: @td333)

4. Great Divide beer dinner @ Porter Beer Bar on 8/19 - \$45 for 3 courses + beer pairings. <http://trunc.it/172wt>

5. Interesting Concept-Bear Petitions.com launches&hopes 2help craft beer drinkers enjoy beer they want @their fave pubs. <http://bit.ly/11gJQN>

6. Beer Cheddar Soup: Dish number two in my famed beer dinner series is Beer Cheddar Soup. I hadn't had too.. <http://bit.ly/1diDdF>

7. New York City Beer Events - Beer Tasting - New York Beer Festivals - New York Craft Beer <http://is.gd/39kXj> #beer

8. Love beer? Our member is trying to build up a new beer drinker's forum. Grab a #beer and join us: <http://tr.im/pD1n>

9. #Baltimore Beer Week continues w/ a beer brkfst, beer pioneers luncheon, drink & donate event, beer tastings & more. <http://ping.fm/VyTgw>

10. Seattle and Beer: I went to Seattle last weekend. It was my friend's stag - he likes beer - we drank beer.. <http://tinyurl.com/cpb4n9>

RESULTSET A (Left)

1. Team is just off to be 'usability tested' I wonder if they will be usable on their return?

2. Camping outside !

3. 10 Questions with Eric Wathen of Alameda Brewhouse. <http://tinyurl.com/58rf8k>

4. Ooh I snagged a treasury! <http://tinyurl.com/6j2gog> woot woot!

5. Internships INTERNSHIP PICK OF THE WEEK The Virginian-Pilot Summer Internship Program: The .. <http://tinyurl.com/5jb7y8>

6. The new Mozilla social browser "Flock" is awesome ! <http://flock.com/>

7. But, Really, Which Palin Is Best Now? <http://tinyurl.com/676cos>

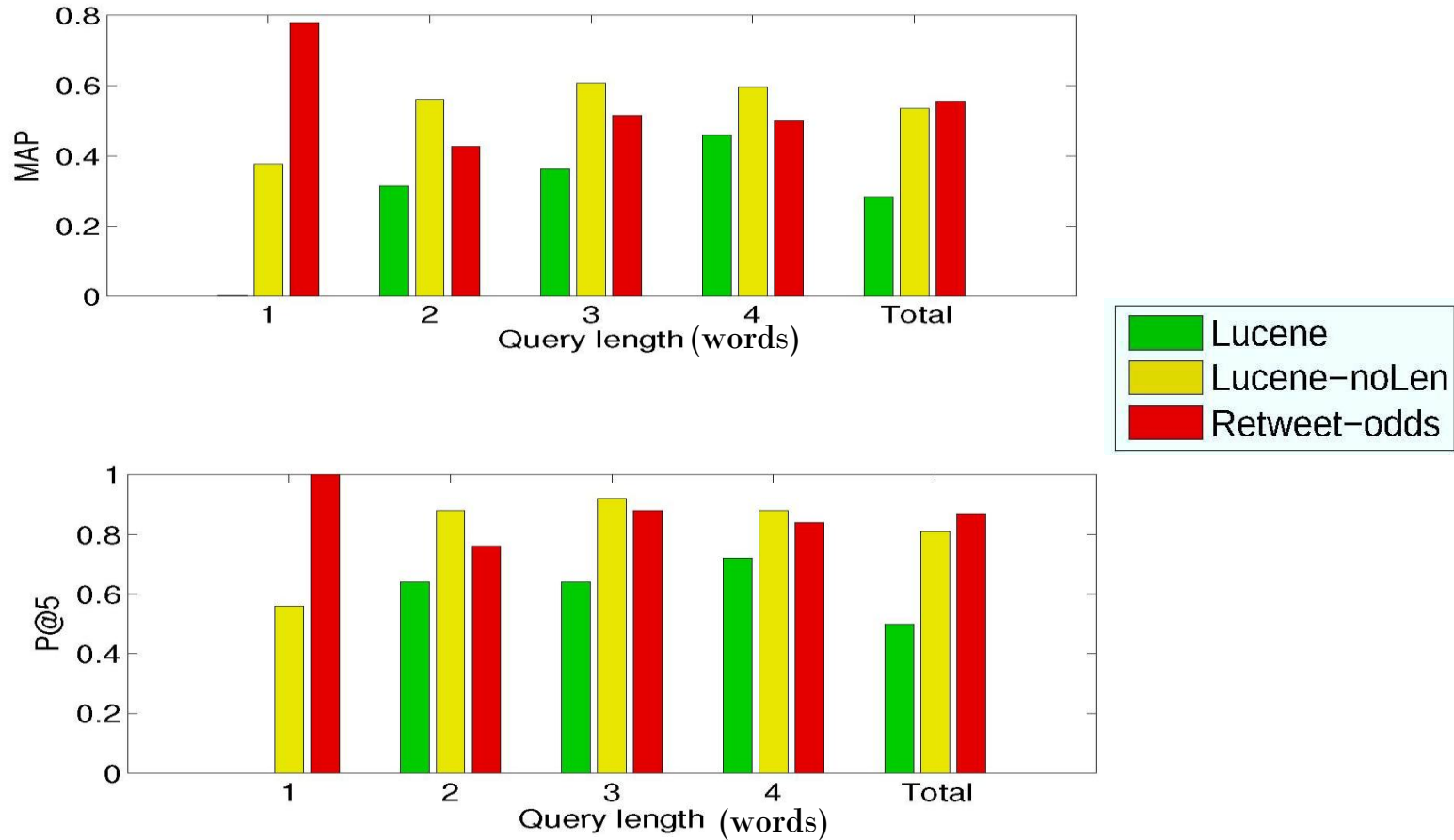
8. "A WORLD WITHOUT WALLS" (TED TURNER) INK5\*\*\*\*\*

9. just found a new blog called convention insider -<http://www.conventioninsider.com/>

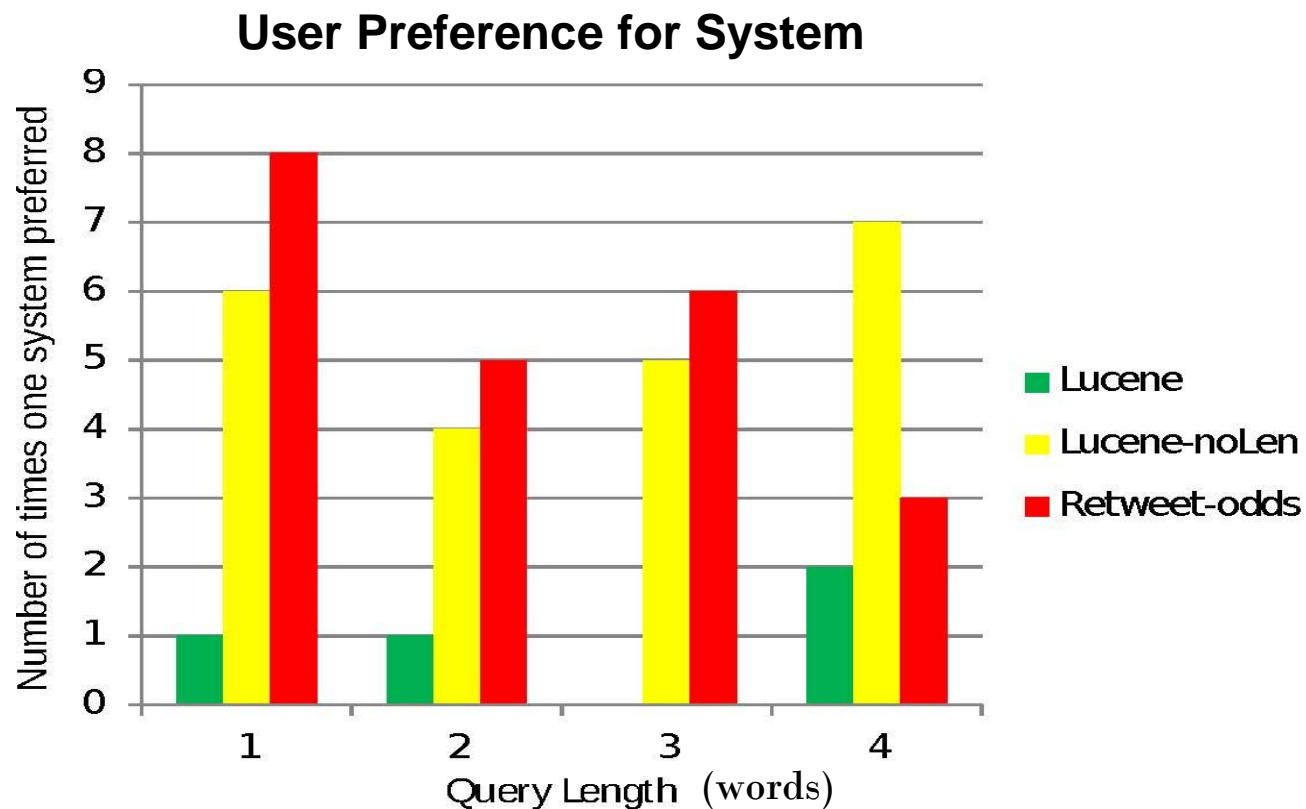
10. Took me forever to write a Haiku... I might've done a better job, but AdWords couldn't wait (blame Google)

RESULTSET B (Right)

## Retrieval Performance



## Retrieval Performance



- **Query:** *beer*

Rank	Retweet Prediction Model	Baseline Model
1	UK beer mag declares "the end of beer writing." @StanHieronymus says not so in the US. <a href="http://bit.ly/424HRQ">http://bit.ly/424HRQ</a> #beer	beer
2	beer summit @bspward @jhinderaker no one had billy beer? heehee #narm - beer summit @bspward @jhinde <a href="http://tinyurl.com/n29oxj">http://tinyurl.com/n29oxj</a>	beer!!
3	Go green and turn those empty beer bottles into recycled beer glasses!   <a href="http://bit.ly/2src7F">http://bit.ly/2src7F</a> #beer #recycle (via: @td333)	BEER
4	Great Divide beer dinner @ Porter Beer Bar on 8/19 - \$45 for 3 courses + beer pairings. <a href="http://trunc.it/172wt">http://trunc.it/172wt</a>	Beer
5	Interesting Concept-Beer Petitions.com launches&hopes 2help craft beer drinkers enjoy beer they want @their fave pubs. <a href="http://bit.ly/11gJQN">http://bit.ly/11gJQN</a>	beer.....

- Problems in Microblog Retrieval
  - ◆ Effect of length normalization
  - ◆ How to measure content quality
  
- Our Contribution
  - ◆ Length Normalization is counter productive
  - ◆ Retweets-odds as measure of content quality and interestingness
  - ◆ Content based retweet prediction model

# Thank You